BELLABEAT - CASE STUDY

WARDA RAEES

12/16/2021

Introduction

Bellabeat is the go-to wellness brand for women with an ecosystem of products and services focused on women's health. The company develops wearables and accompanying products that monitor biometric and lifestyle data to help women better understand how their bodies work and make healthier choices. Bellabeat web-page.

Bellabeat app : The Bellabeat app provides users with health data related to their activity, sleep, stress, menstrual cycle, and mindfulness habits. This data can help users better understand their current habits and make healthy decisions. The Bellabeat app connects to their line of smart wellness products.

The mission statement

Bellabeat, a high-tech company that manufactures health-focused smart products wants to analyse the usage of one of their products in order to gain insight into how people are already using their smart devices.Then, using this information, she would like high-level recommendations for how these trends can inform Bellabeat marketing strategy.

PHASE 1 : ASK

Identify the business task: The company better target their marketing efforts into their customer's needs based on their usage of their fitness smart devices. With this info then make high-level recommendations for how these trends can inform Bellabeat marketing strategy.

Consider key stakeholders: The main stakeholders here are Urška Sršen, Bellabeat's co-founder and Chief Creative Officer; Sando Mur, Mathematician and Bellabeat's cofounder; And the rest of the Bellabeat marketing analytics team.

Questions for the analysis:

- What are some trends in smart device usage?
- How could these trends apply to Bellabeat customers?
- How could these trends help influence Bellabeat marketing strategy?

PHASE 2 : PREPARE AND PROCESS

- I prepare and process the data using SQL on BigQuery and then analyze filtered datasets using R on RStudio.

- The three tables I analyzed from dataset are dailyActivity_merged, sleepDay_merged, and weight-LogInfo_merged, renamed Activity_Filtered, Sleeplog_Filtered, and WeightlogInfo, respectively.

- Filtered datasets and Queries are uploaded here on Github.

PHASE 3 : VISUALIZING DATA

```
library(tidyverse)
```

LOADING PACKAGES

```
## -- Attaching packages -------------------------------------- tidyverse 1.3.1 --
## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.6     v dplyr   1.0.7
## v tidyr   1.1.4     v stringr 1.4.0
## v readr   2.1.1     v forcats 0.5.1
## -- Conflicts ----------------------------------------- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(lubridate)
```

```
##
## Attaching package: 'lubridate'
## The following objects are masked from 'package:base':
##
##     date, intersect, setdiff, union
```

```
library(dplyr)
library(ggplot2)
library(tidyr)
```

```
Activity_Filtered<-read.csv("/cloud/project/BELLABEAT/Fitabase Data 4.12.16-5.12.16/Activity_Filtered.cs
Sleeplog_Filtered<-read.csv("/cloud/project/BELLABEAT/Fitabase Data 4.12.16-5.12.16/Sleeplog_Filtered.cs
weightLogInfo<-read.csv("/cloud/project/BELLABEAT/Fitabase Data 4.12.16-5.12.16/weightLogInfo.csv")
```

```
# explore num of active minutes per category
Activity_Filtered %>%
  select(VeryActiveMinutes, FairlyActiveMinutes, LightlyActiveMinutes) %>%
  summary()
```

IMPORTING DATASETS

```
##  VeryActiveMinutes FairlyActiveMinutes LightlyActiveMinutes
##  Min.   :  0.00    Min.   :  0.00      Min.   :  0.0
##  1st Qu.:  0.00    1st Qu.:  0.00      1st Qu.:146.5
##  Median :  7.00    Median :  8.00      Median :208.0
##  Mean   : 23.02    Mean   : 14.78      Mean   :210.0
##  3rd Qu.: 35.00    3rd Qu.: 21.00      3rd Qu.:272.0
##  Max.   :210.00    Max.   :143.00      Max.   :518.0
```

```
# sleep
Sleeplog_Filtered %>%
  select(TotalSleepRecords, TotalMinutesAsleep, TotalTimeInBed) %>%
  summary()
```

```
##  TotalSleepRecords TotalMinutesAsleep TotalTimeInBed
##  Min.   :1.00      Min.   : 58.0      Min.   : 61.0
```

```
##   1st Qu.:1.00    1st Qu.:361.0    1st Qu.:403.8
##   Median :1.00    Median :432.5    Median :463.0
##   Mean   :1.12    Mean   :419.2    Mean   :458.5
##   3rd Qu.:1.00    3rd Qu.:490.0    3rd Qu.:526.0
##   Max.   :3.00    Max.   :796.0    Max.   :961.0
```
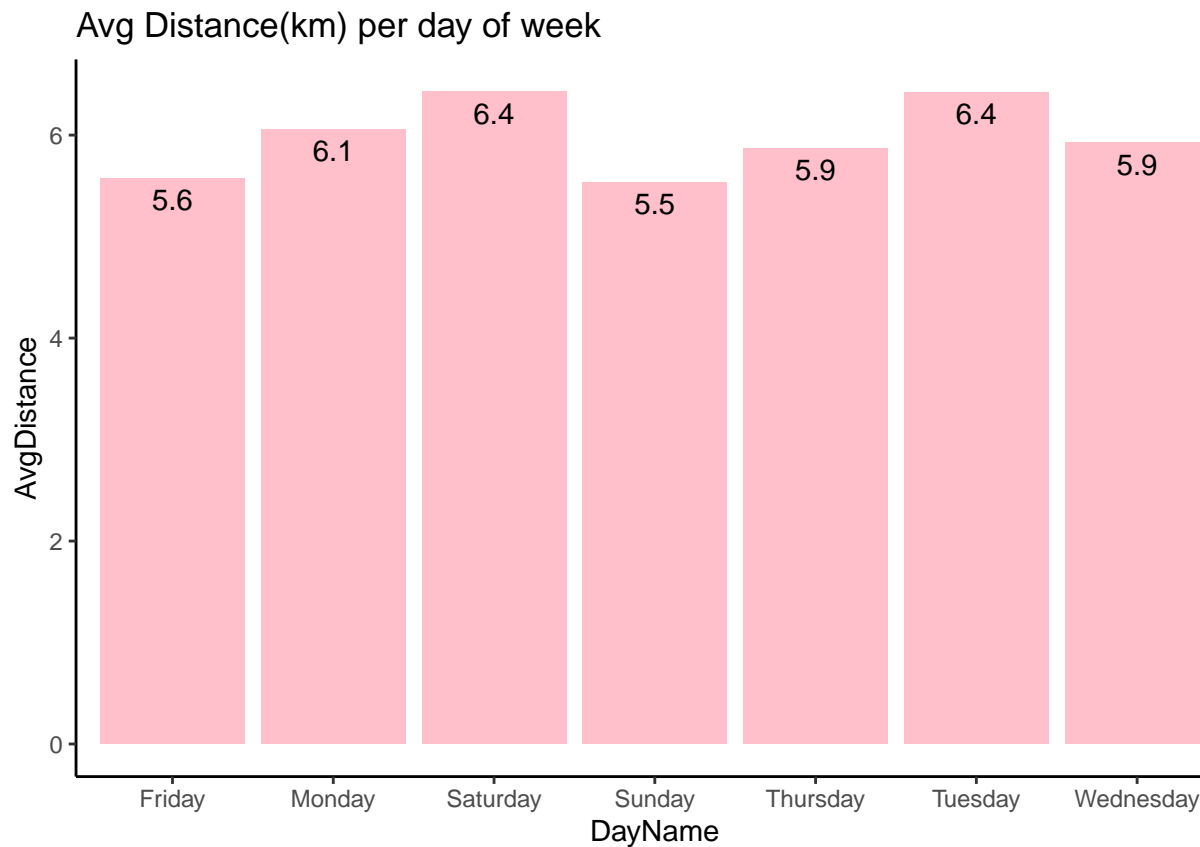
```r
# weight
weightLogInfo %>%
  select(WeightKg, BMI) %>%
  summary()
```

```
##     WeightKg          BMI
##   Min.   : 52.60   Min.   :21.45
##   1st Qu.: 61.40   1st Qu.:23.96
##   Median : 62.50   Median :24.39
##   Mean   : 72.04   Mean   :25.19
##   3rd Qu.: 85.05   3rd Qu.:25.56
##   Max.   :133.50   Max.   :47.54
```

Some Discoveries from the Summaries

- Average sedentary time is 991 minutes or 16 hours.Definately needs to be reduced! *FairlyActiveMinutes < VeryActiveMinutes< LightlyActiveMinutes
- The majority of the participants are lightly active.
- Total Minutes Asleep are less than Total Time In Bed indicating people does not immediately sleep just after going to the bed.
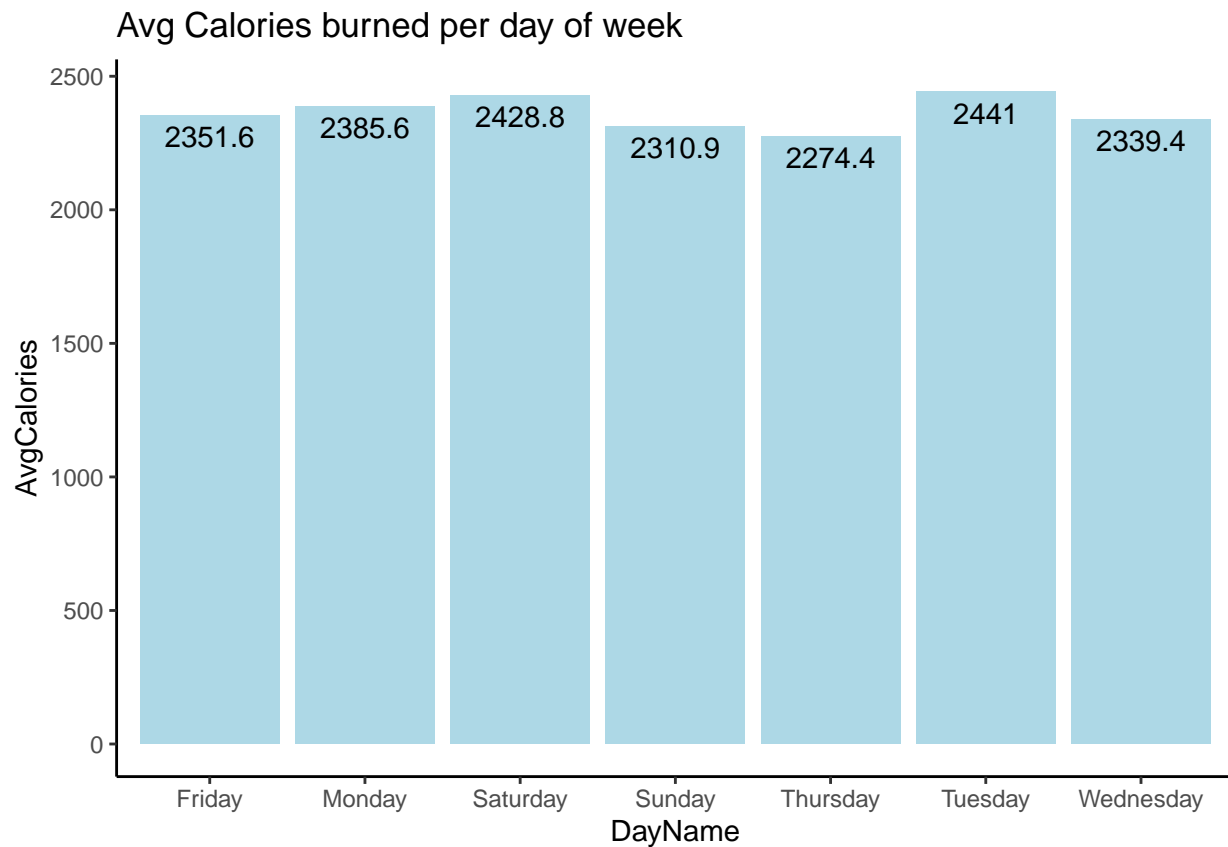
```r
activity<-read.csv("/cloud/project/BELLABEAT/Fitabase Data 4.12.16-5.12.16/activity_av_perday.csv")
ggplot(activity, aes( y=AvgDistance, x=DayName)) +
  geom_bar(position='dodge', stat='identity',fill="pink")+
  geom_text(aes(label = round(AvgDistance, 1)),vjust = 1.5,color = "black")+
  theme_classic()+
  labs(title="Avg Distance(km) per day of week")
```

Avg Distance(km) per day of week

|	Friday	Monday	Saturday	Sunday	Thursday	Tuesday	Wednesday
	5.6	6.1	6.4	5.5	5.9	6.4	5.9

Activity Analysis

Lowest average distance occurred on the weekend-sunday indicating that people mostly don't usually go on morning walks on off days and prefer rest.

```
ggplot(activity, aes( y=AvgCalories, x=DayName)) +
  geom_bar(position='dodge', stat='identity',fill="lightblue")+
  theme_classic()+
  labs(title="Avg Calories burned per day of week")+
  geom_text(aes(label = round(AvgCalories, 1)),vjust = 1.5,color = "black")
```

Avg Calories burned per day of week

		2428.8			2441	
2351.6	2385.6		2310.9	2274.4		2339.4

Friday | Monday | Saturday | Sunday | Thursday | Tuesday | Wednesday

DayName

(y-axis: AvgCalories, with values 0, 500, 1000, 1500, 2000, 2500)

Overall, average calorie expenditure didn't vary much by day of the week.

```
ggplot(activity, aes( y=AvgSteps, x=DayName)) +
  geom_bar(position='dodge', stat='identity',fill="seagreen")+
  labs(title="Avg Steps taken per day of week")+
  geom_text(aes(label = round(AvgSteps, 1)),vjust = 1.5,color = "white")+    theme_classic()
```

Avg Steps taken per day of week
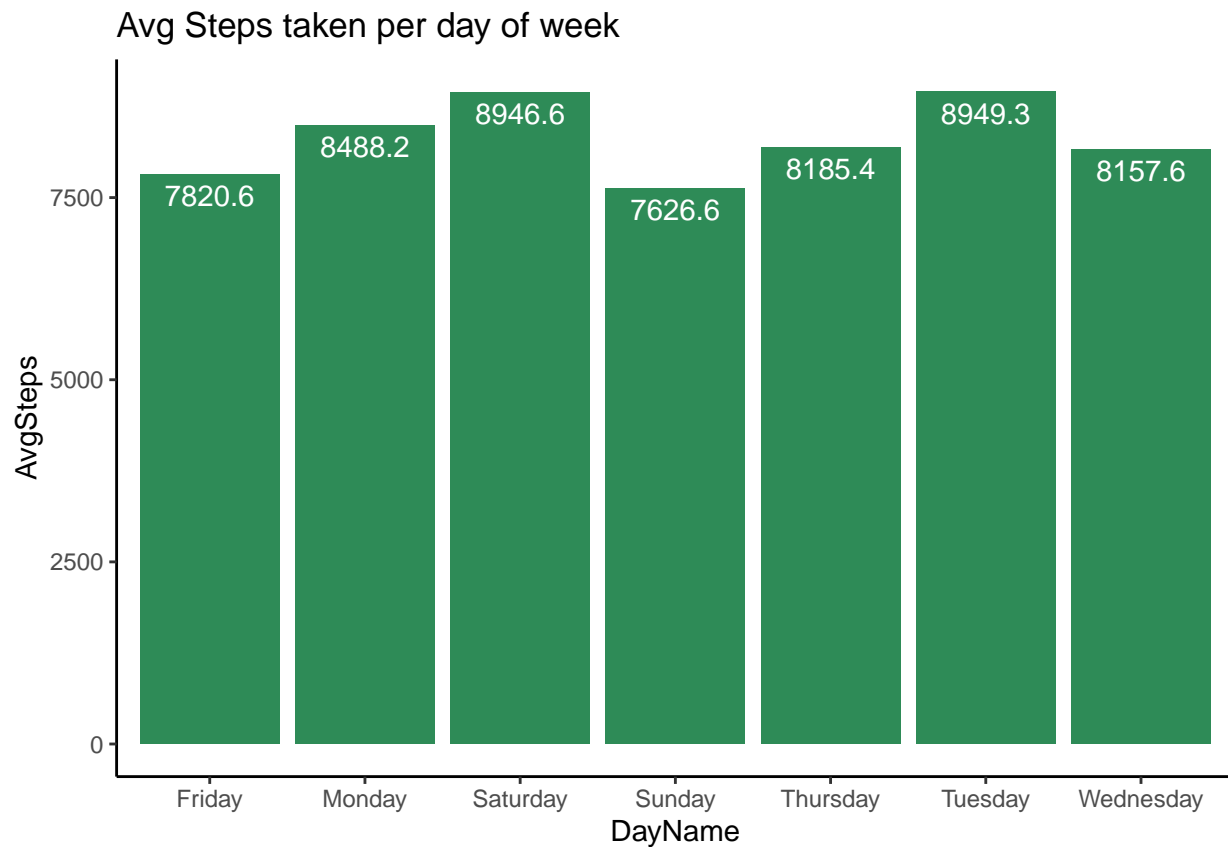
Average daily steps was under the recommended amount of 10,000 steps for every day of the week

```
c1<-read.csv("/cloud/project/BELLABEAT/Fitabase Data 4.12.16-5.12.16/caloriesVmin.csv")

ggplot(data=c1,mapping=aes(x=Calories, y=SedentaryMinutes)) +
  geom_jitter(colour="brown", size=1)+
  geom_smooth(mapping=aes(x=Calories, y=SedentaryMinutes) ,method="gam")+
  labs(title="Calories burned VS Sedentary Minutes")
```
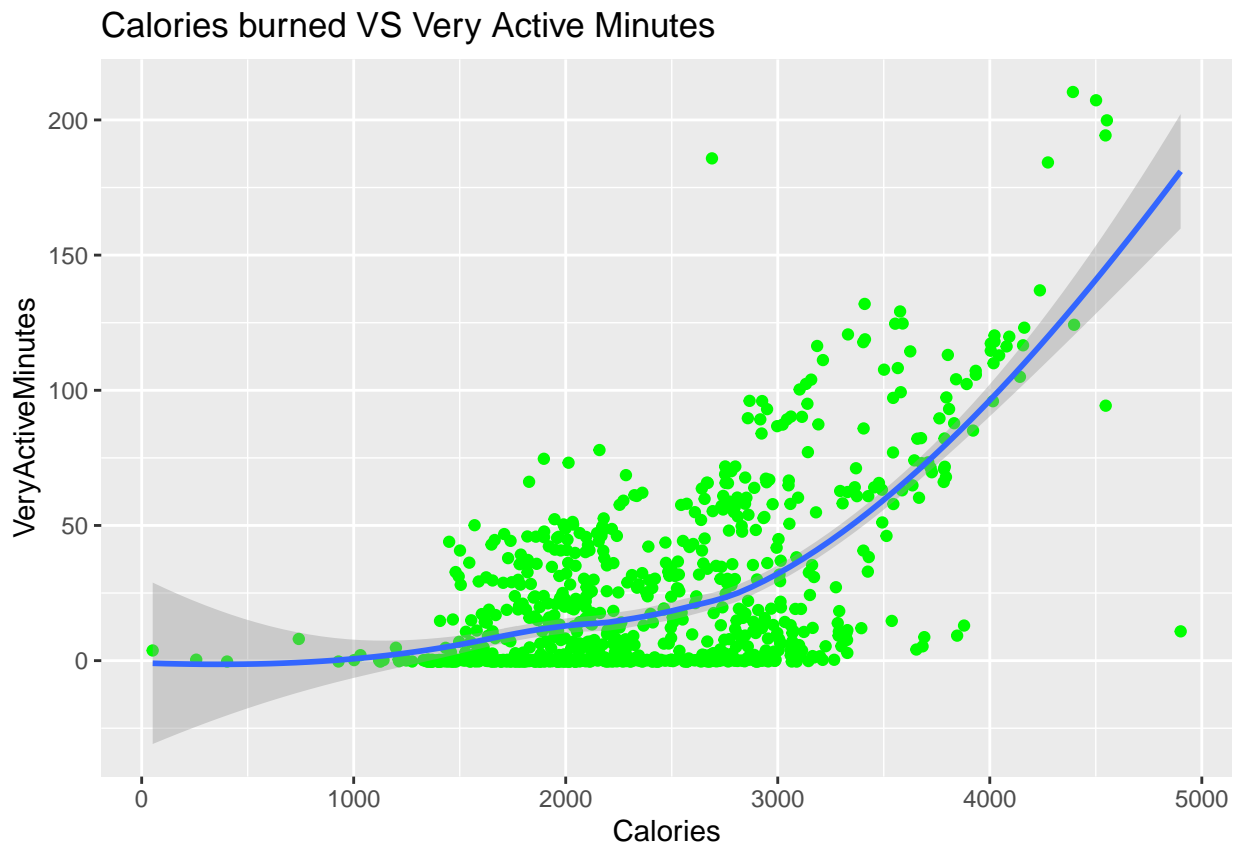
Calories Burned Analysis

```
## `geom_smooth()` using formula 'y ~ s(x, bs = "cs")'
```

Calories burned VS Sedentary Minutes
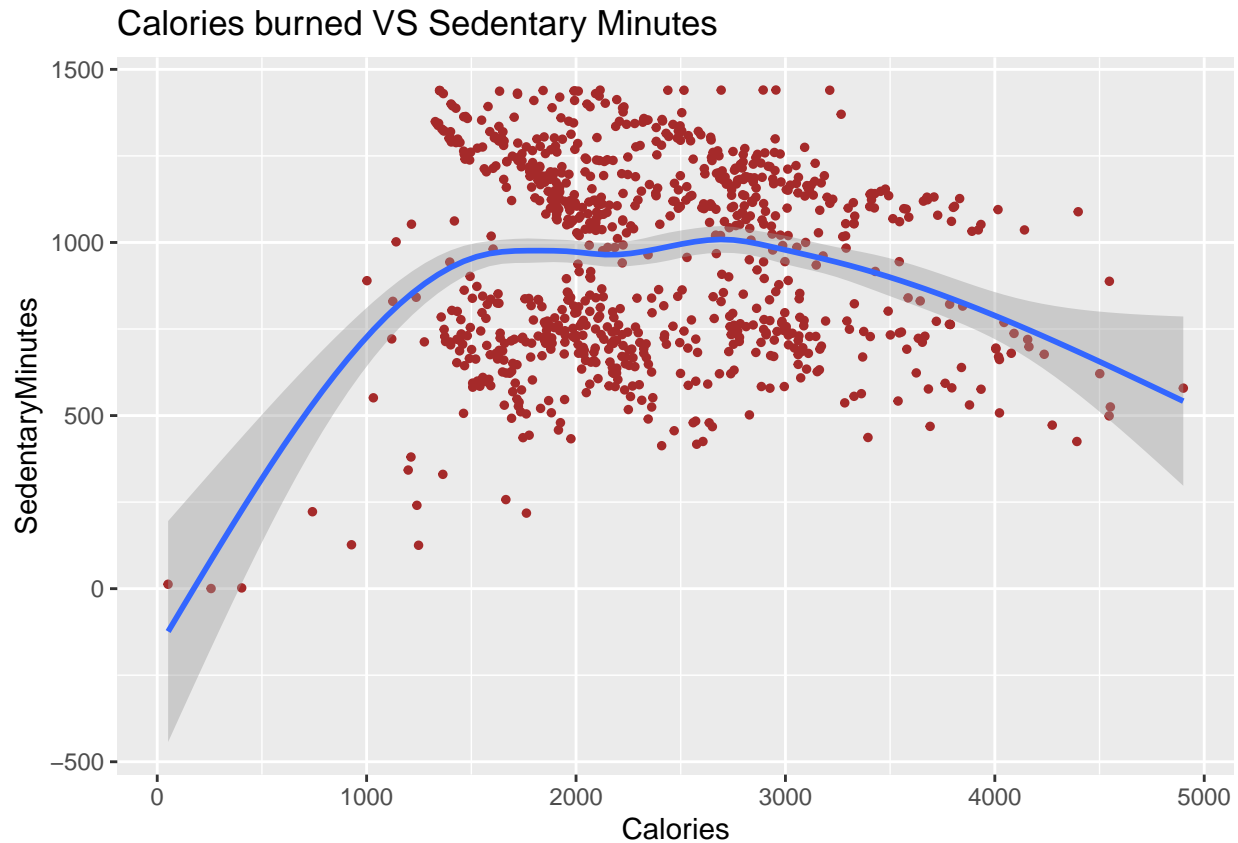
There is no actual correlation here between inactive minutes and Calories, indicating if we want to burn calories then we should be active in our daily routine, refer to the next plot.

```
ggplot(data=c1,mapping=aes(x=Calories, y=VeryActiveMinutes)) +
  geom_jitter(colour="green", size=1.5)+
  geom_smooth(mapping=aes(x=Calories, y=VeryActiveMinutes) ,method="loess")+
  labs(title="Calories burned VS Very Active Minutes")
```

```
## `geom_smooth()` using formula 'y ~ x'
```

Calories burned VS Very Active Minutes

There is somehow positive correlation here between active minutes of participants and Calories, which is obvious - the more active we are, the more calories we burn.
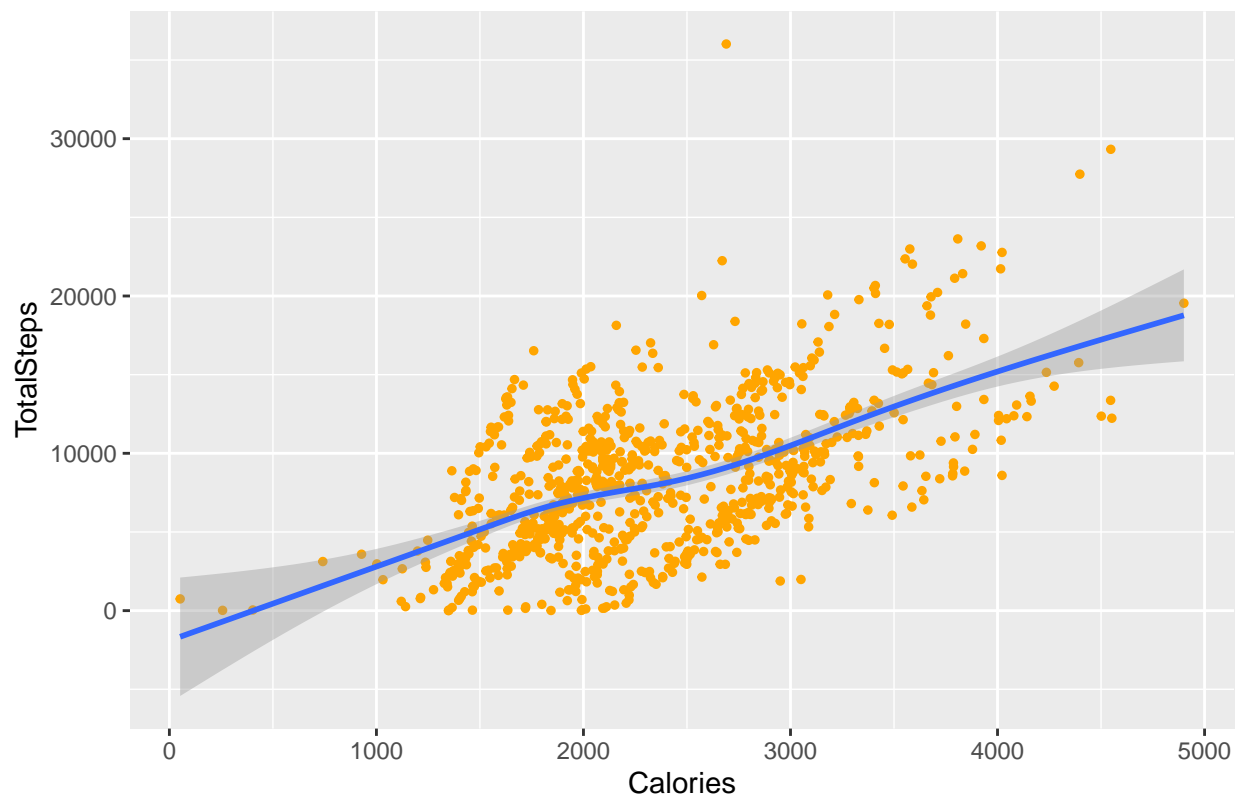
```
c2<-read.csv("/cloud/project/BELLABEAT/Fitabase Data 4.12.16-5.12.16/caloriesVsteps.csv")

ggplot(data=c2, mapping=aes(x=Calories, y=TotalSteps)) +
  geom_jitter(colour="orange", size=1)+
  geom_smooth(mapping=aes(x=Calories, y=TotalSteps) ,method="gam")+
  labs(title="Calories burned VS Total Steps")
```

```
## `geom_smooth()` using formula 'y ~ s(x, bs = "cs")'
```

Calories burned VS Total Steps

There is a positive correlation here between total steps taken by participants and Calories.

Weight Analysis

```r
weight<-read.csv("/cloud/project/BELLABEAT/Fitabase Data 4.12.16-5.12.16/weight_av.csv")

ggplot(weight, aes( x=participants, y=IsManualReport)) +
  geom_bar(position='dodge', stat='identity',fill="grey")+
  labs(title="Participant's Weight Records", subtitle="Data entered is Manually or Automatically")+
  theme_classic()
```

Participant's Weight Records
Data entered is Manually or Automatically
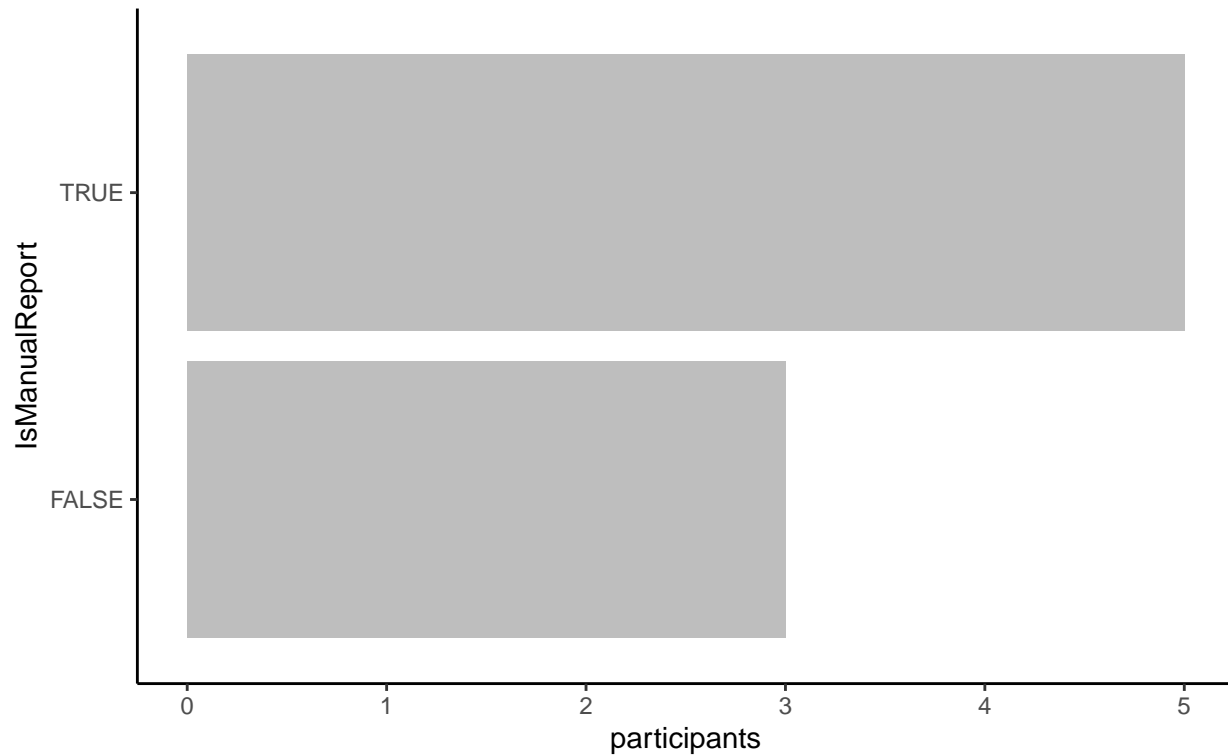
Participants recording weight data manually are more than automatically indicating that people are not confident enough to check their weight automatically by fitbit app.

```
ggplot(weight, aes( x=IsManualReport, y=avg_weight)) +
  geom_bar(position='dodge', stat='identity',fill="tomato")+
  geom_text(aes(label = round(avg_weight, 1)),vjust = 1.5,color = "white")+   labs(title="AVG Weight of
  theme_classic()
```

AVG Weight of Participants w.r.t type of data entered
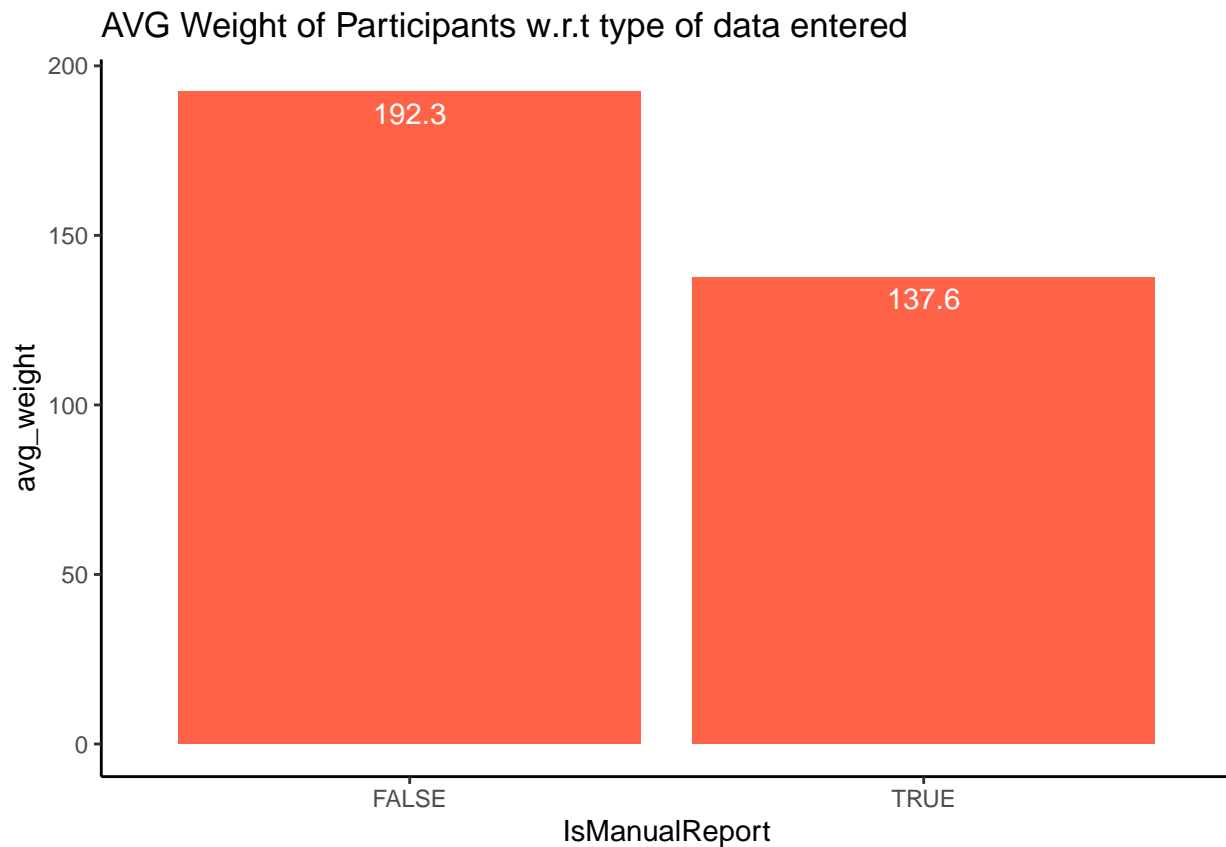
Above hypothesis related of weight consciousness is somehow getting true here, by seing the above plot we can conclude that weight data collected manually is less than automatically- people may enter their false weight.
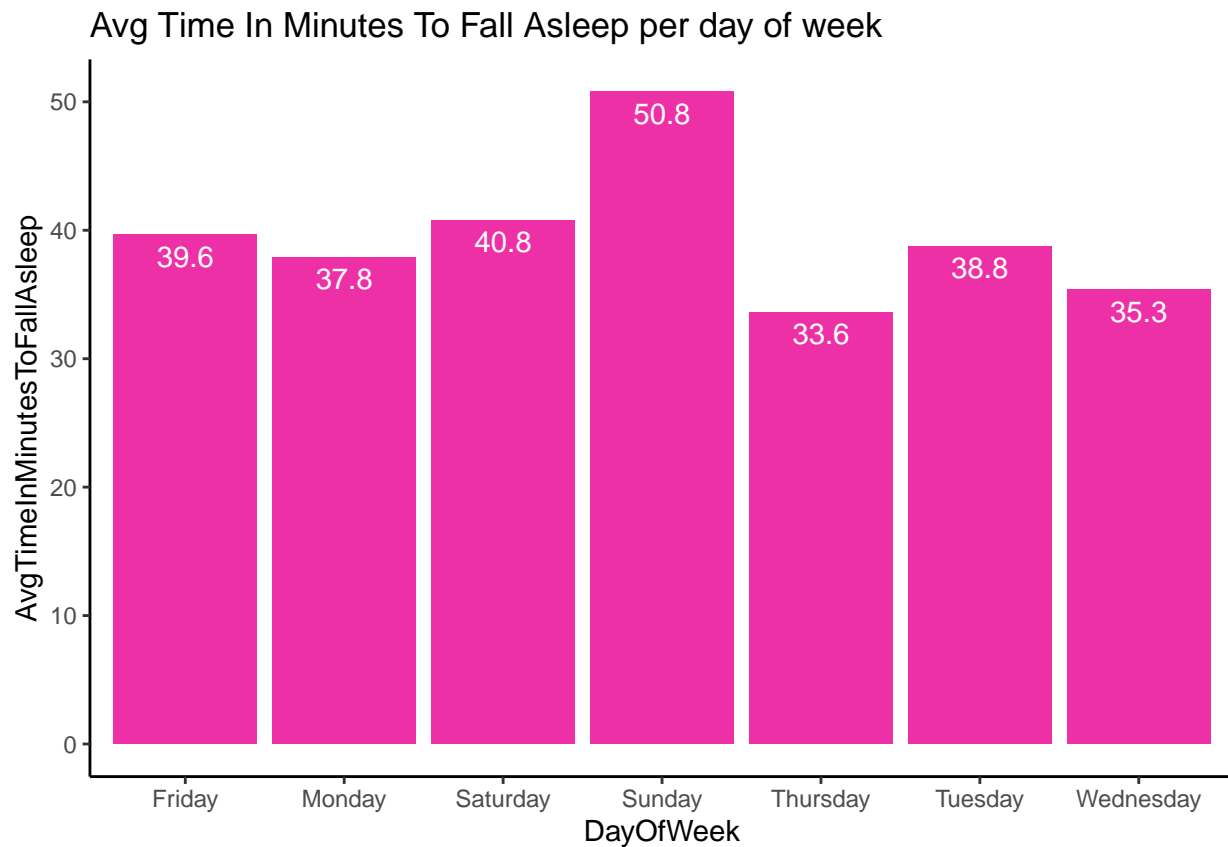
Sleep Analysis

```
sleep<-read.csv("/cloud/project/BELLABEAT/Fitabase Data 4.12.16-5.12.16/sleep_av.csv")

ggplot(sleep, aes( y=AvgHoursAsleep, x=DayOfWeek)) +
  geom_bar(position='dodge', stat='identity',fill="aquamarine1")+
  labs(title="Avg HOURS of sleep per day of week")+
  geom_text(aes(label = round(AvgHoursAsleep, 1)),vjust = 1.5,color = "black")+
  theme_classic()
```

Avg HOURS of sleep per day of week
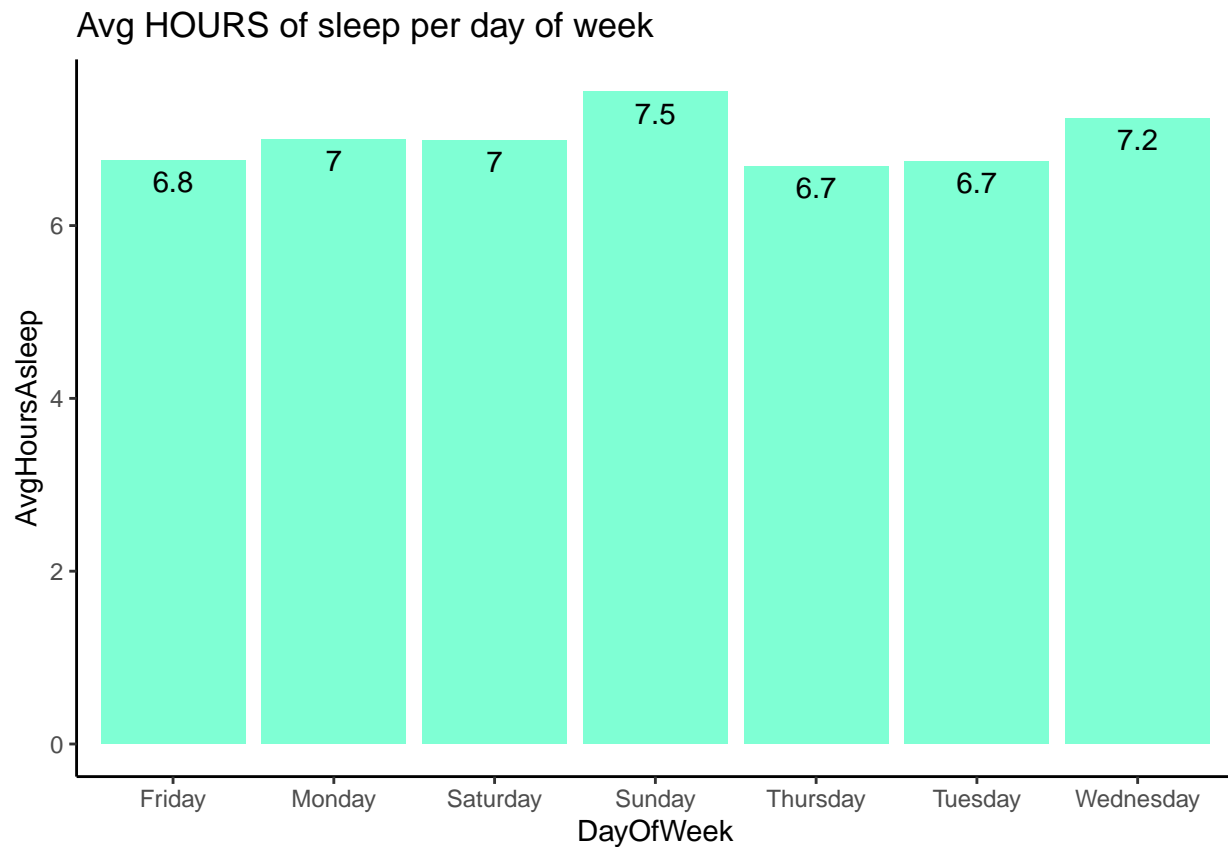
Recommended sleep per night is 7-9 hours or more according to the CDC, but here we can see that on weekdays due to work sleeping hours of parcipants are mostly less than 7.

```
ggplot(sleep, aes( y=AvgTimeInMinutesToFallAsleep, x=DayOfWeek)) +
  geom_bar(position='dodge', stat='identity',fill="maroon2")+
  labs(title="Avg Time In Minutes To Fall Asleep per day of week")+
  geom_text(aes(label = round(AvgTimeInMinutesToFallAsleep, 1)),vjust = 1.5,color = "white")+
  theme_classic()
```

Avg Time In Minutes To Fall Asleep per day of week

From the above plots it is clear that Avg Time takento fall asleep on sunday and saturday is highest because usually participants spent the most time awake in bed on Sunday nights.

```r
sleep2<-read.csv("/cloud/project/BELLABEAT/Fitabase Data 4.12.16-5.12.16/sleep_absent.csv")

ggplot(sleep2, aes( y=users_Absent_sleepdata, x=DayName)) +
  geom_bar(position='dodge', stat='identity',fill="lightgreen")+
  labs(title="No. of missing sleep records per day of week")+
  geom_text(aes(label = round(users_Absent_sleepdata, 1)),vjust = 1.5,color = "black")+
  theme_classic()
```

No. of missing sleep records per day of week
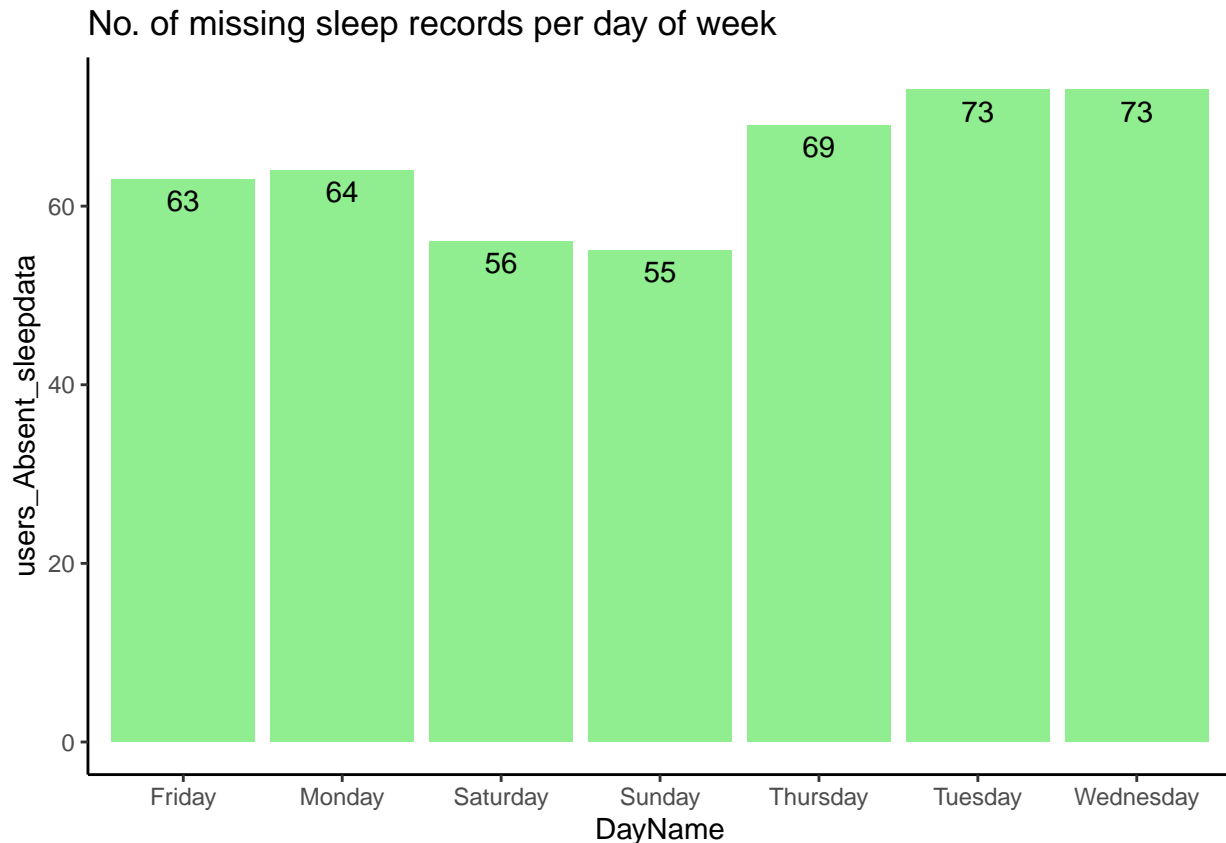
Above plot reveals that their is a huge No.of missing sleep records on each day of week.

PHASE 4 : SHARE AND ACT

ANALYSIS OF THE FITBIT APPS DATA ANSWER THE FOLLOWING QUESTIONS OF USERS

- How active are my days? Do I spend a considerable amount of time being sedentary?
- How does this data vary on weekdays vs weekends?
- What factors contribute to the highest calorie burn?
- Which exercises are the best and easiest way to achieve good health?
- Have I been following a steady sleep schedule? What factors influence it?
- What is the impact of a Netflix binge on weekend sleep?

CONCLUSIONS

- Participants in this study collected far more activity data than data on sleep or weight.
- On average, participants could've used more sleep throughout the week.
- One way to do this might've been decreasing sedentary activity (by increasing the amount of time spent active). *Positive correlation between steps taken and calories is a highlight, encouriging being active in order to remain healthy.
- Sleep was less frequently recorded on weekdays, but there were a lot of missing sleep records in general. Similarly, a quarter of the participants didn't record any sleep data at all. Improvement in both of these areas would increase the reliability of the data.
- If users want to burn calories, they're best off doing so by being very active and/or increasing their daily step count.
- Giving any statement on the basis of weight data is not feasible with the current dataset as only 8 parcipants recorded their weight data, reasons could not be highlighted correctly with currently available

dataset.

RECOMMENDATION

- if users want to improve their sleep, the Bellabeat app can recommend reducing sedentary time and Bellabeat app should give out notifications to go to bed.
- Train a simple Machine Learning model to see if there is a hidden pattern to attain better sleep.
- Market the Bellabeat products as being comfortable and light weight so that user can wear it even when he/she is sleeping, as their is less sleeping data.
- I will recommend to design a marketing strategy that will help their users to produce more data by starting research on Bellabeat customers tracking habits.